The FHWA Travel Model Improvement Program Workshop over the Web

The Travel Model Development Series: Part I – Travel Model Estimation

presented by
Thomas Rossi
Cambridge Systematics, Inc.

November 6, 2008

1

Webinar Outline

- Session 1: Introduction – October 16, 2008
- Session 2: Data Set Preparation – November 6, 2008
- Session 3: Estimation of Non-Logit Models – December 11, 2008
- Session 4: Estimation of Logit Models – February 10, 2009

2

Webinar Outline (continued)

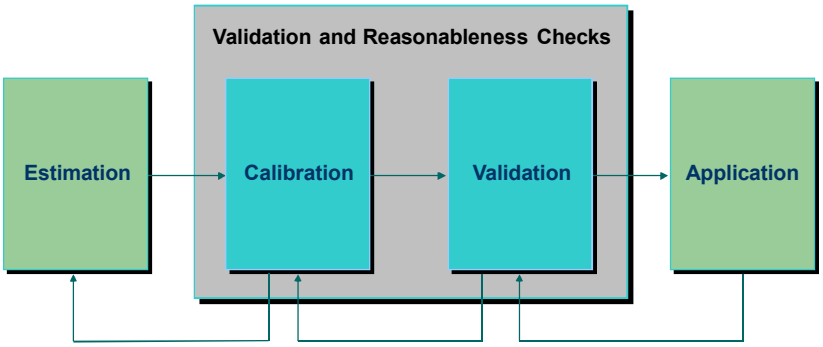
- Session 5: Application and Validation of Logit Models – March 12, 2009
- Session 6: Advanced Topics in Discrete Choice Models – April 14, 2009
- Session 7: Trip Assignment – May 7, 2009
- Session 8: Evaluation of Validation Results – June 9, 2009

3

Homework
From Session 1

4

The Role of Data in Travel Modeling



The Role of Data in Travel Modeling

- Estimation
  - Local data for parameter estimation
- Validation/calibration
  - Observed data for comparisons and checks
- Application
  - Network, socioeconomic, and other data

6

The Importance of Data Quality

- Potential data quality problems
  - Errors in data collection, computation, transcription, etc.
  - Incorrect data processing
  - Out of date information
  - Statistical insignificance

7

The Importance of Data Quality

- Effects
  - Incorrectly estimated model parameters
  - Incorrect input data (garbage in…)
  - Model application inconsistent with context
  - False precision of results

8

Data for Model Estimation

- Local survey data
- National data (Census, NHTS)
- Network data/skims
- Socioeconomic data
- Other (parking costs, auto operating costs)

9

Data for Model Application

- Network data/skims
- Socioeconomic data
- Other (parking costs, auto operating costs)

10

Data for Model Validation

- Local survey data
- National data (Census, NHTS, NCHRP 365)
- Observed travel information
  - Traffic counts
  - Transit ridership/boardings
  - Highway speeds

11

Model Estimation Data Sources

- Household activity/travel survey (household, trip level)
- Transit on-board survey
- Other surveys
- Critical nonsurvey data
  - Socioeconomic data
  - Networks
  - Other (area types, parking costs, auto operating costs, etc.)

12

Model Types

Household level	• Auto ownership • Trip production
Trip level	• Mode choice • Trip distribution (logit)
Aggregate	• Trip attraction • Trip distribution (gravity) • Time of day

13

Person Data File
From Household Survey

- Each record represents a person
- Each field represents a characteristic of the person (age, gender, worker status, student status, etc.)
- Not used directly in most four-step models (household based)
- Often used in person based models such as activity based
  - Would include characteristics of the household for model estimation

14

Household Data File
From Household Survey

- Each record represents a household
- Each field represents a characteristic of the household or its location

15

Household Data File
Typical Fields

From the survey

- Location (zone/point)
- Number of persons
- Number of workers
- Number of children
- Number of autos
- Income level
- Number of trips by purpose

From other sources

- Area type of zone
- Residential and commercial density
- Accessibility measures (for auto availability)

16

Household Data File
Data Checks

- Completeness (fields, members of household)
- Consistency checks
  - Consistency with person file
  - Number of persons >= # of workers, # of children, etc.
  - Numbers add up, e.g. persons = males + females
- Reasonableness checks
  - Distributions of households by # of persons, # of workers, # of autos, income level, etc.
- Geocoding errors
- Weights
  - Weights should sum to the population represented for each segment

17

Household Data File
Dealing with Missing/Incorrect Data

- Missing data
  - Can they be deduced?
  - Should they be imputed?
  - Income is commonly missing from 10-20% of records (refusals to respond)
  - Add a field or value to indicate missing data
- Incorrect data
  - Can they be corrected? Not usually
  - Survey response or coding errors
  - Failed logic/consistency checks

18

Trip Data File
From Household/On-Board Surveys

- Each record represents a trip made by an individual
- Each field represents a characteristic of:
  - The trip;
  - The traveler;
  - His/her household; or
  - The areas traveled

19

Trip Data File
Typical Fields

From the survey

- Origin and destination
- Trip purpose
- Chosen mode
- Time of day of trip
- *Trip time/cost*
- Household/person characteristics (linked from household/person file)

From other sources

- Travel time (in-vehicle)
- Other time components (wait, access/egress, transfer)
- Costs (parking, auto operating, transit fare)
- Number of transit transfers
- Zone attributes
- Logsums from other models

20

Trip Data File
Why Not Use Reported Level of Service Data?

- Rounding of responses to 5, 15, even 30 minutes
- Perception bias varies among individual respondents
- Need a consistent source of information for all records
- Need information for non-chosen alternatives

21

Trip Data File
Attaching Data from Other Sources

- Index data to be attached based on an identifier in the survey data records (e.g. zone number)
- Set up other data sources as lookup table

Zone	Area Type
1	5
2	4
…	…
n	2

22

Trip Data File
Data Checks

- Logic/reasonableness checks
  - Reported mode consistent with travel time
  - Reported times/costs consistent with skims
  - Chosen mode availability
  - Excessive times/costs/transfers
  - Consistency of times of day for each person
  - Origin of trip = destination of last trip
  - Origin and destination must be different for each trip
  - Bus routes used

23

Combining Household and On-Board Survey Data

- Data not appearing in all surveys
- Differences in question wording
- Differences in data ranges
- Surveys done at different times
- Changes in transportation system

24

Setting up Data for Disaggregate Model Estimation

1. Assemble survey data
2. Data checks
3. Create necessary variables
   a. Maximum values
4. Attach skim data
5. Data checks
6. Designate choice variable

25

Setting up Data for Aggregate Model Estimation
Trip Attraction Model - linear regression

- Define independent variables to be tested
- Use trip file – weighted data
- Aggregate to districts
- Attach district level data
  - Employment by type
  - Households

26

Setting up Data for Aggregate Model Estimation
Gravity Model

- Define independent variable (e.g. highway travel time)
- Use trip file – weighted data
- Compute trip length frequency distribution by trip purpose

27

Setting up Data for Aggregate Model Estimation
Time of Day Model

- Determine resolution for testing (e.g. half hours)
- Use trip file – weighted data
- Define time variable (e.g. departure time, arrival time, midpoint)

28

Homework
Session 2

29